



## Summary

### Contents

#### A Quarterly Information Bulletin from the Methodology Division

##### Contents

Indigenous Analysis Program  
Testing of ABN's on front of ABS Business Forms  
Reintroduction of Trend Estimates for Retail Time Series  
Analysis of FACS Administrative Dataset  
Generalised Regression Variance Estimation for ABS Economic Surveys  
Neighbourhood Income Inequality in Australian Cities  
Form Design Standards refresh for Optical Character Recognition (OCR)  
Small Area Estimates of Crime and the propensity to report crime to Police

## Indigenous Analysis Program

### INDIGENOUS ANALYSIS PROGRAM

Enhancing the statistics about the Indigenous population of Australia is a high priority for the ABS with considerable resources invested in the Indigenous aspects of ABS collections and administrative datasets. Recently, the Analysis Board approved a portfolio of proposals for additional analytical work in this field.

Analysis Branch has already been undertaking a variety of joint projects, such as:

- Developing a suite of experimental indexes that summarise the relative socioeconomic condition of the Indigenous populations in different geographic areas. This project was done jointly with the Western Australia Indigenous Statistics Unit and the National Centre for Aboriginal and Torres Strait Islander Statistics [NCATSIS]. The project team was led by Ruel Abello.
- Analysing the sensitivity of income distribution measures (for the Indigenous and non-Indigenous populations) to choice of unit and to assumptions about the pattern of resource sharing within income units and households. This project is being done jointly with the Centre for Aboriginal Economic Policy Research at the Australian National University. Steven Kennedy is the leader of the ABS project team.
- Assessing the quality of the Indigenous coding of a key administrative databank, namely the Family and Community Services-Centrelink Longitudinal Dataset. This is of interest both to those who use the databank for policy design and assessment and to the ABS which might wish to use the databank to compile Indigenous statistics. This project is led by Ruel Abello.

Future topics may include:

- Indigenous demography. ABS demographers are advising Dan Smith about the choice of a topic.
- Administrative datasets. Several projects might explore ways of using Indigenous-coded administrative datasets to augment the data from ABS collections. We may trial this idea using administrative datasets available in Western Australia.
- Regional estimates. There is considerable interest in using survey and administrative data to synthesise Indigenous estimates for finer geographic areas than are available at present. We may establish a pilot project to explore the methodological and data issues.

- Estimating Indigenous "service populations". This project would try estimate the number and characteristics of Indigenous people who receive or may be eligible for government health, education, labour market and other services.

Our preferred approach is to establish joint project teams with ABS specialists in Indigenous statistics and experts in subject matter fields (e.g. demography, health, labour market, etc.) The teams would report to an Indigenous Analysis Project Board which has members from Social Conditions Statistics Branch, NCATSIS and other areas.

For more information, please contact Ken Tallis on (02) 6252 7290.

E-mail: [ken.tallis@abs.gov.au](mailto:ken.tallis@abs.gov.au)

## Testing of ABN's on Front of ABS Business Forms

### TESTING OF ABN'S ON FRONT OF ABS BUSINESS FORMS

The Australian Business Number (ABN) system has recently been implemented into Australian businesses as part of Federal Governments' reform of taxation. There is a subsequent push for government agencies to use ABNs in their transactions with the public. These factors have led the ABS to consider how we could use ABNs to our advantage as an identifier for linking statistical units and as a Sample and Frame Maintenance trigger. A standard approach for confirming ABNs on questionnaires was needed and testing was initiated by Economics Standards Section to find the best way to do this. The Forms Consultancy Group's (FCG) interest in the project is in the process of testing the effectiveness of the chosen questions and what impact the questions may have on the layout of the rest of the questionnaire.

#### The BUS test

The Birthing Units Survey (BUS) was selected as the vehicle for the first round of testing. A split sample design allowed the comparison of two different form types, each asking the respondent to confirm if a preprinted ABN was the one that belonged to their business. One form type used a "please correct" box above the address and the other used two direct questions. The aim of the test was to demonstrate if one form type was more effective than the other or if neither was effective. The form types would be ineffective if respondents currently had ABNs that were different from the preprinted number on their form but did not change it. There is a high risk of this happening if respondents do not read the questions properly.

#### The Post Enumeration Survey (PES)

To evaluate how effective the two form types were a telephone PES was conducted on a subsample of respondents. FCG developed interview scripts for the PES, with assistance from the Victorian Maths Stats Cell (MSC) and Economics Standards, to ensure consistency between interviews. Four different scripts had to be written to cover four respondent groups: those who received each of the two form types, by those who changed their ABN and those who didn't. Different scripts, some sequencing and an infinite number of potential responses from respondents made the task quite complicated! The PES confirmed the business's ABN, and sought explanations of any differences. It was tested by FCG and the MSC on businesses who had responded to the BUS and based on testing the script was altered. The full PES was then passed on to the Business Frames Section to conduct.

#### The Error Analysis

To complement the PES with a slightly more quantitative test, FCG designed an error analysis for the two form types. This involved figuring out all the different ways respondents could have obviously filled out the ABN questions incorrectly or ignored them altogether. The Vic MSC then selected a subsample of returned forms and physically examined them to count the proportion of these errors. This examination turned out to be extremely valuable in assessing one of the form types and unlike the PES, it was not time-critical.

#### The Way Forward

While there were some limitations to the BUS test, we obtained some important information that we expect to feed into the development of improved ABN questions for a further test, preferably on a subannual survey. We also learned a lot about the process of testing changes to forms with several different groups of stakeholders and this has already fed into the planning for the next round. We intend to conduct similar procedures as in the last test, but administrative data from the ATO are also being investigated in the hope that this will provide supplementary information or enable more effective targeting of respondents.

For more information, including the actual results of our testing, please contact Emma Farrell on (02) 6252 7316.

## Reintroduction of Trend Estimates for Retail Time Series

### REINTRODUCTION OF TREND ESTIMATES FOR RETAIL TIME SERIES

A trend estimate describes the underlying behaviour of a statistical time series by excluding the impact of seasonal and irregular, or unusual, influences on the data. The New Tax System (TNTS), which introduced the Goods and Service Tax (GST) and removed some Wholesale Sales Taxes from 1 July 2000, has both short and long term impacts on Australian retailing. The Australian Retail Trade series is recorded inclusive of the GST because it is used in the national accounts, and more generally, to measure household final consumption expenditure (i.e. the actual value paid by consumers). Short term impacts associated with the introduction of TNTS are not considered part of the underlying behaviour of the series.

As the TNTS impact could not be accurately measured at the time, the Australian Bureau of Statistics decided to suspend the trend estimate from July 2000 until spending behaviour stabilised and the impacts could be accurately measured and removed. As expected, it took some months for regular spending patterns to be re-established and this process was further complicated by unusual spending associated with the Sydney 2000 Olympic Games.

The introduction of TNTS impacted on the Retail Trade series in four ways:

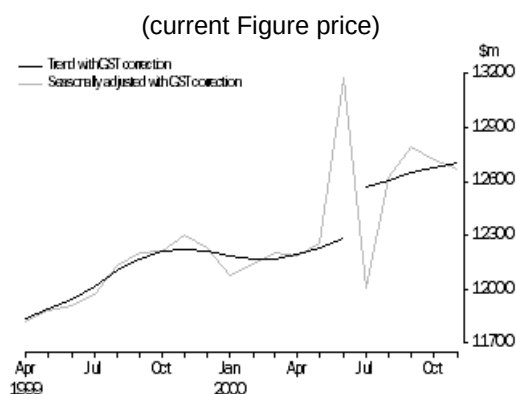
- spending patterns changed prior to the introduction of the GST. The principal impact on retail turnover pre-GST was an increase in the volume of goods sold as consumers brought purchases forward to June 2000.
- from 1 July 2000 the Retail Trade series recorded turnover inclusive of the GST. The net effect of the TNTS (i.e. the removal of the Wholesale Sales Tax and the inclusion of the GST) resulted in a change in the valuation basis of the series, so a permanent level shift was introduced.
- spending patterns in July 2000 were unusual as many industries had a decrease in sales as a result of consumers having brought purchases forward to June. While the price of most goods rose from 1st July 2000, the volume of goods sold across most industries and States dropped.
- spending patterns were affected for a number of months after July (but not as dramatically as for the July month), and stabilised spending patterns emerged slowly and differentially across industries.

Time Series Analysis section undertook special corrections to stabilise the seasonally adjusted estimate from June to November 2000 and monitored whether the spending behaviour had stabilised. The longer the period of time the trend was suppressed, the more chance there was that non-GST related effects would affect the estimation of the level shift size at July 2000. Balancing the competing demands, the ABS chose to use data up to November 2000 as a good compromise to reintroduce trend estimates.

An intervention analysis was used to simultaneously estimate the impact on the Australian retail trade series of unusual spending in June and July 2000 (treated as prior corrections) and a permanent level shift (trend break) between June and July due to the inclusion of the GST, and exclude the Olympic impact in September.

Figure 1 shows the trend and seasonally adjusted estimates of the total Australian retail turnover after the pre-GST and post-GST impacts were appropriately prior corrected. The estimated level shift provides a starting point of the trend under the TNTS environment.

Figure 1: Australian total retail turnover



The Retail Trade current price trend series was reintroduced with the December 2000 issue of the ABS's Retail Trade publication (cat. 8501.0). The impact of unusual pre-GST and post-GST (not a permanent level shift starting from July) spending was removed from the chain volume series and the series provided a good measure of the levels and changes in the volume of retail turnover.

For more information, please contact Mark Zhang: (02) 6252 5132

E-mail: [mark.zhang@abs.gov.au](mailto:mark.zhang@abs.gov.au)

## Analysis of FACS Administrative Dataset

### ANALYSIS OF FACS ADMINISTRATIVE DATASET

The Analysis Branch now has access to an administrative dataset that contains (nameless) information about persons receiving welfare payments from Centrelink. Through a Memorandum of Agreement with the Department of Family and Community Services (FaCS), the Branch is able to look at an extensive range of data about social security customers in the form of a longitudinal data set (LDS). The dataset tracks fortnightly payments to social security customers and captures customer characteristics, three quarters of which are changing over time.

The wealth of information contained in the LDS provides a comprehensive picture of flows into, out of, and transfers within the welfare payments system. It helps to capture the dynamic nature of welfare, and the characteristics of customers that affect these dynamics.

A one percent sample of the LDS data has been copied to a SAS data set and has been made available to the Branch for analysis.

ABS has a special interest in administrative and business by-product data. The Corporate Plan 2000 states that an expanded and improved national statistical service can be achieved by, among other strategies, 'better utilisation of both public and private administrative and transactional data sources'. This include "developing methods for utilising the very large but imperfect datasets available through administrative and transaction data holdings...".

This year, with this broad objective in mind, the Branch started to use the data to pursue a couple of research topics. The aim is to give both ABS and FaCS a better understanding of the strengths and weaknesses of this very large administrative dataset, whose characteristics differ greatly from traditional censuses and surveys. The analysis focuses on aspects like data quality, useability, interpretability, and linkability with other data sources. It is hoped that in the long run this administrative dataset could offer opportunities to create new analytical or statistical products that may serve as replacements for, or supplements to, data that ABS currently collects directly.

The two research investigations (also called modules) that are currently underway are:

- Module 1: Duration, transition and flow analyses of disability support and mature age customers; and
- Module 2: Analysis of LDS metadata, including an investigation of the quality of Indigenous coding and implications for analyses of Indigenous issues.

The first module intends to help FaCS understand better the dynamics of welfare participation of certain groups of customers (people with disability, mature age customers, Indigenous people). The analyses make use of a variety of methods, including a suite of techniques called survival analysis. The project team is developing and testing various non-parametric, parametric and semi-parametric models to estimate survivor and hazard functions and the influences of covariates (e.g. demographic characteristics) on the probabilities of customers entering into a particular payment program, exiting it, or transferring to a different one.

The second module aims to make recommendations for improving the data structures and descriptions of the LDS, for research purposes. It looks into the quality of Indigenous identification and the degree of useability of the LDS to studies that have Indigenous perspectives.

For more information, please contact the team leader, Ruel Abello on (02) 6252 5511. The other members of the team are Anil Kumar, and John de Maio.

E-mail: [ruel.abello@abs.gov.au](mailto:ruel.abello@abs.gov.au)

# Generalised Regression Variance Estimation for ABS Economic Surveys

## GENERALISED REGRESSION VARIANCE ESTIMATION FOR ABS ECONOMIC SURVEYS

The Australian Bureau of Statistics (ABS) is investigating improving the efficiency of its estimation methodology for its economic surveys by using Business Activity Statements (BAS) that are submitted to the Australian Taxation Office by Australian businesses. Improving efficiency will allow the ABS to reduce the sample size of its surveys, and consequently, the cost of producing publication estimates, while maintaining their level of accuracy. Statistical agencies around the world are facing similar challenges, as administrative data are becoming accessible.

General Regression (GREG) is a statistical estimation methodology that may use several items of auxiliary information, such as the various items available in BAS data. GREG estimation is potentially significantly more efficient than the alternatives, which are to use either no auxiliary information or a single auxiliary item as in ratio estimation. The ABS has an estimation system, called GREGWT, that implements the GREG estimation methodology.

While the ABS can calculate GREG estimates for its economic surveys by using GREGWT, it must also be able to estimate the precision, or variance, of these estimates. Special Projects Unit in the Statistical Services Branch is drafting a working paper evaluating methodologies for estimating GREG variances, including: Bootstrap, Jackknife and Balanced Repeated Replication. This evaluation includes theoretical criteria as well as the ability to meet ABS' output requirements.

The working paper recommends the Bootstrap for calculating GREG variance estimates for ABS economic surveys. It develops the theory supporting the use of Bootstrap to calculate variances of level and rate estimates and their movements between two time points, in single-phase and two-phase sampling schemes.

For more information, please contact James Chipperfield on (02) 6252 7301.

E-mail: [james.chipperfield@abs.gov.au](mailto:james.chipperfield@abs.gov.au)

## Neighbourhood Income Inequality in Australian Cities

### NEIGHBOURHOOD INCOME INEQUALITY IN AUSTRALIAN CITIES

This project is tackling two questions about the distribution of income in Australian cities:

- Are neighbourhoods within cities becoming more income-homogeneous?
- How are changes in income inequality within and between neighbourhoods related to changes in city-wide inequality?

Typically, studies of income inequality describe changes in population income inequality; they examine particular **population sub-groups** that characterise (or dominate) segments of the income distribution, but the **spatial** distribution of income is rarely examined. Describing changes in income inequality in Australian cities will provide information about how cities are developing and whether there are emerging geographical patterns of disadvantage.

The project team will use data from the Census of Population and Housing covering the period 1976 to 1996.

Before we can begin the analysis, we have to consider what defines a "neighbourhood". The most disaggregated level of geography available in Australian Census data is Census Districts (CDs). In most cities, CDs contain on average 200 households and 600 people. We plan to use CDs (as well as a higher level of geography) to characterise neighbourhoods.

We will address a number of additional issues such as:

- how to estimate income inequality indices when we use grouped rather than continuous income data;
- taking account of differences in household size and composition when measuring individual income; and
- describing how the distribution of education and employment are related to the distribution of income.

Statistics Canada recently completed a project that analysed changes in income inequality in Canadian cities. Once the ABS has completed its analysis, the ABS and Statistics Canada plan to compare results. Given the many

similarities in institutional arrangements and urban and economic development between Australia and Canada, these comparisons will help identify explanations for changes in income inequality.

The Project Team is Nick Biddle, Steven Kennedy and Cristy Williams.

For more information, please contact Steven Kennedy: (02) 6252 5462.

E-mail: [steven.kennedy@abs.gov.au](mailto:steven.kennedy@abs.gov.au)

## **Form Design Standards Refresh for Optical Character Recognition (OCR)**

### **FORM DESIGN STANDARDS REFRESH FOR OPTICAL CHARACTER RECOGNITION (OCR)**

A number of changes relating to data capture will occur over the next few years. At the highest level these will entail offering providers a range of integrated multi-modal reporting in many collections, including paper, telephone and electronic reporting in its various forms. For the most common (even traditional) of these, namely paper forms, there is likely to be a major push to use the central scanning facilities operated by the Forms Handling Unit (FHU) for initial data capture as well as form addressing, despatch and receipt.

As survey areas who use or have converted to OCR will be aware, the OCR system currently available for ABS collections implies a number of form design or layout requirements in addition to those specified for paper forms. In particular, precision is necessary in the placing of some design elements, certain colours that are acceptable for paper forms are not suitable for OCR forms, and data capture fields (text and numeric data entry boxes) need to be designed to accommodate the OCR equipment.

As OCR technology is being used for an increasing number of forms with more varied design needs, additional design and structural aspects are constantly coming to light. To date these have largely been handled on an adhoc basis following advice from the FHU. The Forms Design Manual sections relating to the design of OCR forms, as well as the related range of objects and examples available in the standard Pagemaker libraries and templates have not kept up with the increasing scope and demand for OCR form elements. The Forms Consultancy Group (FCG) and FHU are jointly reviewing the standards and range of objects available with a view to updating them.

The 'refreshed' OCR standards will represent an updating and aggregation of existing FCG and FHU standards and practice on OCR design standards, with a major source being existing OCR forms that already include suitable solutions.

One advantage which will follow from this approach will be that form designers will save time using a standard set of Pagemaker objects because they will not have to invent, test or 'borrow' appropriate graphics and the objects used should meet FHU and scanning needs.

With two exceptions, the design features of existing OCR forms that are already in production are unlikely to be affected, and the aim will be to bring them up to date when major changes are made. The two exceptions relate to:

- standardising the size and position of the 'front of form' address label to obtain savings on postage and envelopes; and
- a minor wording change to completion instructions to only ask providers to use black pen.

A draft of the updated standards and graphic objects will be circulated for comment in June, initially through the Economic Statistics Group Collection Strategy and Content Workgroup (CSCW). After comments have been received and any problems worked through they will be incorporated into the 'published' Forms Design Manual for use from September quarter onwards.

For more information, please contact Robert Burnside on (02) 6252 7816.

E-mail: [rob.burnside@abs.gov.au](mailto:rob.burnside@abs.gov.au)

## **Small Area Estimates of Crime and the propensity to report crime to Police**

## SMALL AREA ESTIMATES OF CRIME AND THE PROPENSITY TO REPORT CRIME TO POLICE

These projects are part of a set of projects being done jointly with the National Centre for Crime and Justice Statistics (NCCJS) and the Australian Institute of Criminology (AIC). They use the 1998 National Crime and Safety Survey to:

- develop small area estimates of crime; and
- estimate the propensity of different groups to report crime to the Police.

Both projects use regression methods to analyse the 1998 National Crime and Safety Survey (1998 NC&SS) unit record file.

The "small area estimates of crime" model estimates a regression associating crime rates from the 1998 NC&SS, with socio-demographic data from the 1996 Census. This model is estimated for statistical regions. The coefficients derived are then applied to Statistical Local Areas (SLA), to derive estimates of crime by SLA. The assumption implicit in this is that the characteristics of crime-prone communities at the statistical region are the same as they are at the SLA level. This technique is called synthetic estimation in the literature.

The "propensity to report crime" model uses logistic regression to estimate the propensity to report crime to the Police. The Yes/No reporting behaviour is modelled against victim characteristics from the 1998 NC&SS. The model is estimated using the 1998 NC&SS unit record file.

The outcome of the projects is of particular interest for criminologists and practitioners in the justice field. In particular, there has been considerable interest from Police and the Departments of Justice in each State.

Both models will be validated using sensitivity analysis, and, where definitions have not changed much, the 1993 National Crime and Safety Survey.

Some additional issues are being addressed in this work, such as:

- how to apply regression models to complicated sample designs; and
- how to use small area estimation techniques to get reliable estimates for small areas without increasing the sample size.

We are taking into the account sample design in a number of ways, and will compare the different results. The two broad ways we are taking sample design into account is using the survey weights in some way, and using replicate weights.

The project team is Robert Tanton and René Jones.

For more information, please contact Robert Tanton on (02) 6252 5506

E-mail: [robert.tanton@abs.gov.au](mailto:robert.tanton@abs.gov.au)

## About this release

Methodological News

---

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence